



**Development Research Group,
Development Economics,
World Bank**

ADePT: Labor

Version 1.0

**Automated analysis of the labor market
conditions in low- and middle-income
countries**

Technical User's Guide

The ADePT software is created in the Poverty Team of the Development Research Group, Development Economics SVP by the group lead by Michael Lokshin, Senior Economist, and included Zurab Sajaia and Sergiy Radyakin. The project was completed under guidance of Martin Ravallion.

ADePT Labor is an integrated set of programs that allows users to produce tables needed for analysis of labor market conditions in low- and middle-income countries.

The current version of ADePT Labor processes individual level data and generates about 30 tables on employment status, earnings, working hours, inequality, as well as poverty and child labor. The documentation for ADePT Labor program consists of the following documents:

- “A Guide for Assessing Labor Market Conditions in Developing Countries”
- "ADEPT Labor. User's Guide: Definition of Variables and Indicators"
- “ADEPT Labor. Technical User’s Guide” (this document)
- “Getting Started with ADePT Labor” (a shorter version of the “Technical User’s Guide”)

The set of tables that ADePT Labor produces corresponds to the guidelines defined in “A Guide for Assessing Labor Market Conditions in Developing Countries”. The program also produces a report with basic statistics on all variables and a list of errors, warnings and notes about the variables used in the analysis. While ADePT is operated from within Stata¹ and written in Stata programming language, ADePT itself is window-based and the user does not need to know Stata to work with the program.



We believe ADePT could be a valuable tool for the analysis of labor market conditions. However we warn users that ADePT is not a substitute for good data and for good understanding of the economic principles of labor markets.

ADePT provides limited indications on the quality of the data specified by the user, but this is clearly not a substitute for the careful data cleaning and checking. As with any computer program, the saying “garbage in – garbage out” could be applied to ADePT.

Please read carefully the user’s guides listed above before working with ADePT. Many questions users have about ADePT can be answered by reading these manuals, which also contains references on the relevant literature to help users understand economic methods and techniques used in ADePT.

¹ Stata is a statistical package produced by StataCorp LP, 4905 Lakeway Drive, College Station, Texas, 77845, USA. More information can be found on it’s homepage: <http://www.stata.com>

INTRODUCTORY TOPICS

- 1.1 System requirements
- 1.2 Acquiring and installing ADePT
- 1.3 Maintenance
- 1.4 Dataset layout
- 1.5 Program's outline

1.1 System Requirements

Operating system/platform: although ADePT itself is not platform specific and will run on any platform where Stata can be run, it was tested on Microsoft Windows platform with Microsoft Office Excel and OpenOffice Calc products installed to open the resulting tables.

Hardware: The data files that will be processed by the program determine memory requirements. We recommend having free memory at least 2-3 times larger than the largest file to be processed, since Stata loads the complete dataset into memory. Larger amounts of physical memory generally improve performance.

Stata version: ADePT requires Stata 9.2 or later to be installed on the user's computer.

Necessary ADO files: ADePT will check if required modules are present on the user's computer, and if not, or if their versions are not compatible, ADePT will recommend updates for those specific modules.

1.2 Installing ADePT

An internet connection is required for the installation procedure. Once the installation is complete, an internet connection is no longer required. If questioned by your firewall software, please allow Stata to connect to the internet.

To install ADePT on your computer type in the **Stata** command line:

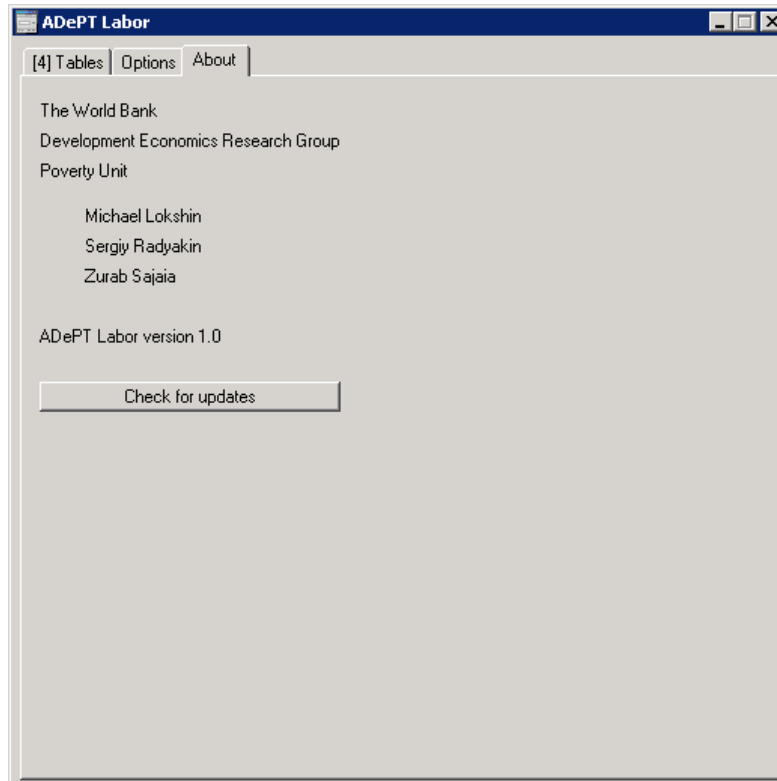
```
net install adept, replace from(http://siteresources.worldbank.org/INTPOVRES/Resources)
```

The above line instructs Stata to download the main module of ADePT. In the first run, it will check for presence and versions of all required modules and advise the user to perform updates if necessary. If all modules have status "up to date" the installation of ADePT completed successfully.

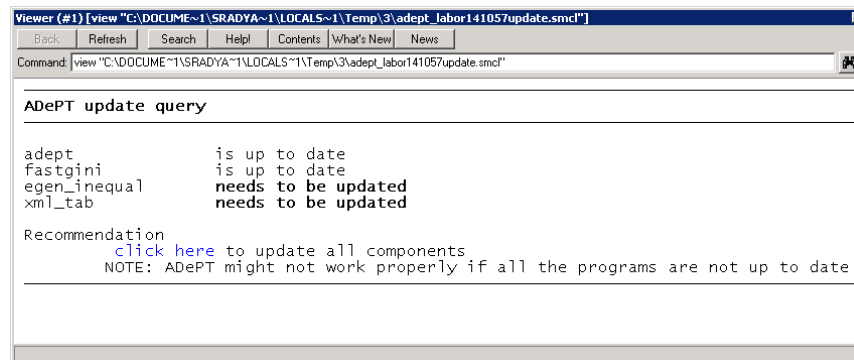
1.3 Maintenance

ADePT has internal integrity checks, which monitor versions of all of the installed components and notify the user if a different version needs to be downloaded.

To check if ADePT itself requires an update, go to the '**About**' tab and click "**Check for updates**" button.



In a new window the status of each module is displayed. We recommend clicking "**Update all**" as long as any module has a new version available.



1.4 Dataset layout

- ADePT works with an individual file expanded with the household-level variables outlined here:

Household identifier	Household member identifier	Household-level characteristics	Individual-level characteristics
X	1	H _X	I _{X,1}
X	2		I _{X,2}
X	3		I _{X,3}
Y	1	H _Y	I _{Y,1}
Y	2		I _{Y,2}
Y	3		I _{Y,3}
Z	1	H _Z	I _{Z,1}
Z	2		I _{Z,2}
Z	3		I _{Z,3}
Z	4		I _{Z,4}

Notice that the household characteristics are the same for all household members (e.g. H_X is the same for all members of the household X), while individual characteristics may differ between the members of the same household.

- If the dataset is in two files, one needs to merge² the household data from the household file to the individual file to obtain a file of the structure shown above. The following variables can be used:

Household level		Individual level	
N	Household identifier <input type="checkbox"/>	N	Household identifier <input type="checkbox"/>
D	Urban/rural indicator	D	Household head indicator
C	Region	N	*Age
N	Welfare aggregate and poverty line	D	Gender
D	Poverty status	C	Education
N	Weight	C	Ethnicity or race
		C	Relationship to household head
		C	*Economic status
		C	*Work Category
		C,D	Sector of activity, agriculture
		N	*Earnings

N – Continuous variable; C – Categorical variable; D – Dummy variable; * - Required variable

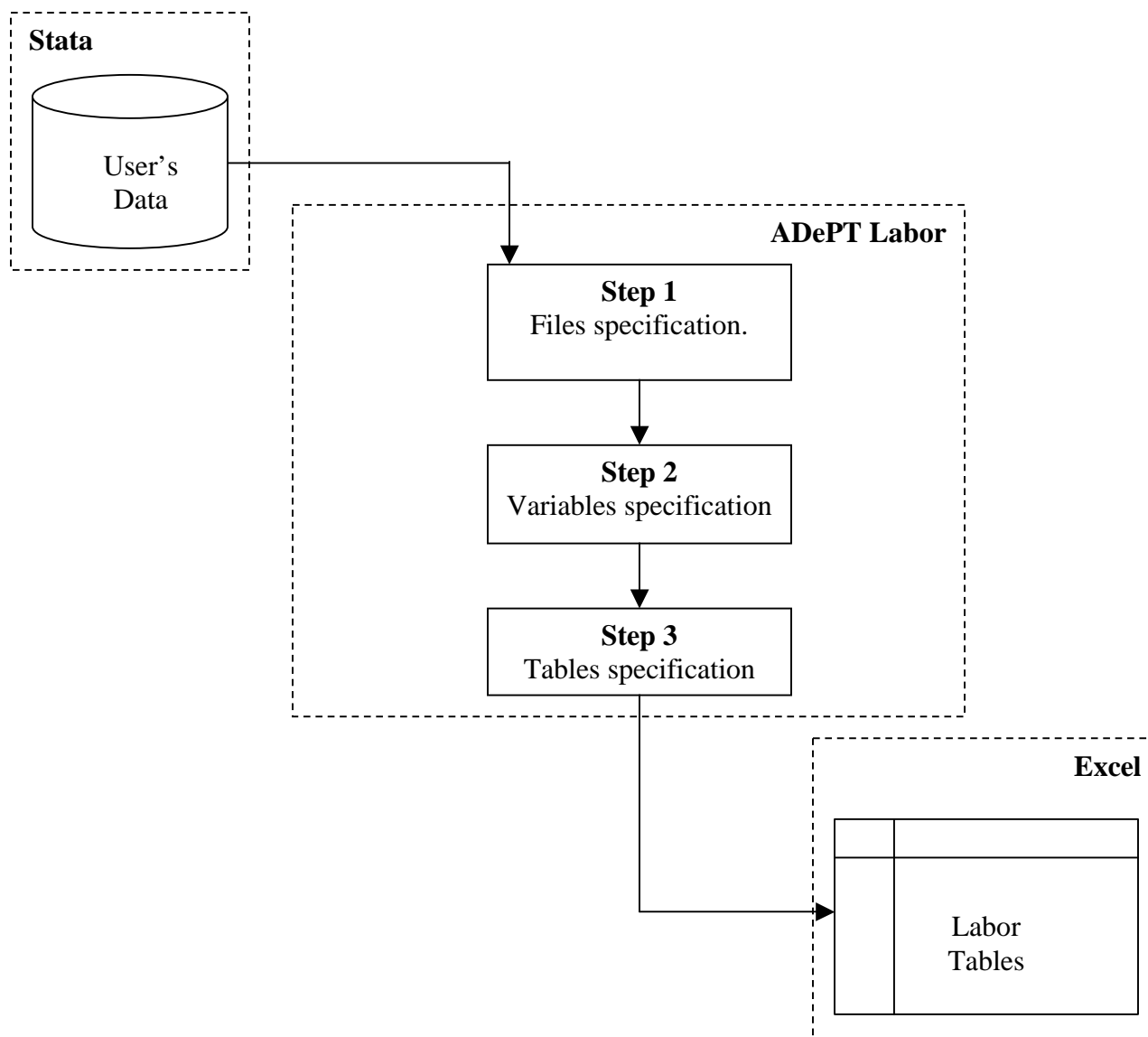
Note that household identifier, the variable that allows merging the information from the two files, must be present in each of them.

² See Stata's manual for description of the "merge" command.

1.5 Program's outline

ADePT incorporates a user-friendly window interface at the input stage and exports the resulting tables in XML files, which can be opened, e.g. by Microsoft Excel on both Windows and MacOs based computers or by Calc program from the OpenOffice on Windows and Linux based computers.

The following diagram represents the steps of producing the tables for labor market conditions analysis using ADePT:



These steps are implemented as a series of screens where the user specifies information regarding files, variables and tables. To accommodate all the variables, step 2 is divided into two screens, other steps take just one screen each.

BASIC TOPICS

- 2.1 ADePT's window, status line and navigation buttons
- 2.2 Step 1. Files specification
- 2.3 Step 2. Variables specification
- 2.4 Step 3. Tables specification
- 2.5 Results

2.1 ADePT's window, status line and navigation buttons

Before launching ADePT make sure that the currently loaded data is saved, because ADePT will change the memory settings and current data will be lost. To start ADePT Labor type in Stata command line:

```
adept labor
```

ADePT program interacts with the user through a series of windows, which correspond to the steps of data input outlined above. All windows consist of the following items:

- current screen's title and number for a quick reference.
- screen-specific controls
- status line
- control buttons:

'Help' displays an online help file, which contains brief information on ADePT's interface and parameters. Refer to the printed manual for detailed explanations.

'Reset' clears all the fields and returns the user to the first screen. The table definitions (described later in this manual) are not cleared by this button. They are automatically cleared after ADePT is restarted.

'Cancel' closes ADePT dialog. Any unsaved changes will be lost.

'<Back' allows to return to the earlier steps of the data input. This button is disabled on step one.

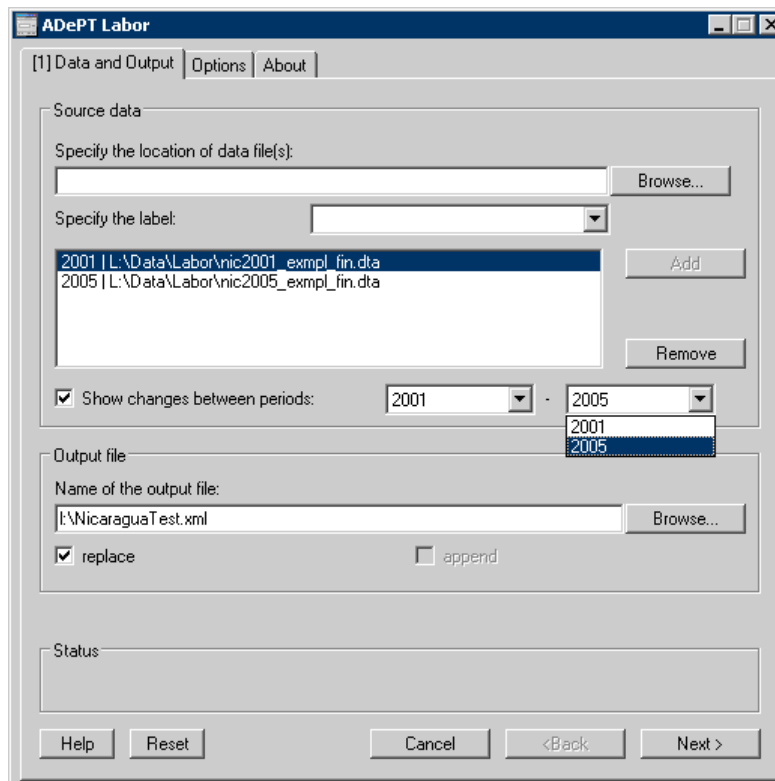
'Next>' advanced forward to the next step of the data input process. This button changes to 'Finish' in the last screen.



Note: ADePT will disable all buttons during long computations. This is a normal behavior. The buttons' states will be restored once the computations are completed.

2.2 Step 1. Files specification

The purpose of the first screen is to specify which files to use in the analysis and where the results should be saved.



To add a file to the list, first specify the filename in the textbox at the top of the form by either entering the filename directly or clicking '**Browse...**' button and selecting a file in the dialog form.

The button '**Add**' remains disabled until a label is assigned in the next field. A label may be a number (e.g. 2005 for a dataset containing data for the year 2005) or a short description. It will be used in the column headings in the resulting tables. Once both filename and label are specified, the '**Add**' button becomes enabled. Click '**Add**' to append the file to the list³.

ADePT works with Stata datasets only. If ADePT can't open a dataset it will notify the user. To add another file to the list, use the '**Browse**' button again, then assign a new label. To remove a file from the filelist, highlight this file by clicking on its name in the list and click '**Remove**'.

When two or more files are selected, the option to compute the changes between the indicators from two datasets becomes enabled. If active (indicated by a checkmark), the two drop-down lists with dataset labels become visible. Choose the first and the second period by clicking on the appropriate labels in these lists.

Under the source data files list the output filename must be specified. A filename can be typed in directly or chosen in a standard file open dialog by clicking a button located near the edit box. The resulting tables will be generated and saved into this file. The output file is produced in XML format,

³ Each file may be listed only once. An attempt to add a file which is already in the list (no matter what the label is) causes an error message

which is supported by the Microsoft Office Excel (on Microsoft Windows and Apple MacOS platforms) and OpenOffice Calc (on Microsoft Windows and Linux platforms).

If an existing file is specified, the program issues a message and refuses to continue. To continue do any of the following:

- provide another filename
- delete this file
- specify an option '*replace*' or '*append*'

The '*replace*' option instructs ADePT to substitute the report specified (the older version will be lost). Option '*append*' tells ADePT to add the tables in the new sheets of the specified file⁴.

If the output file is open by another application (e.g. Excel), the user will be notified. To resolve the situation:

- close an application which is currently using this file (Excel), or
- specify another filename for the report file.

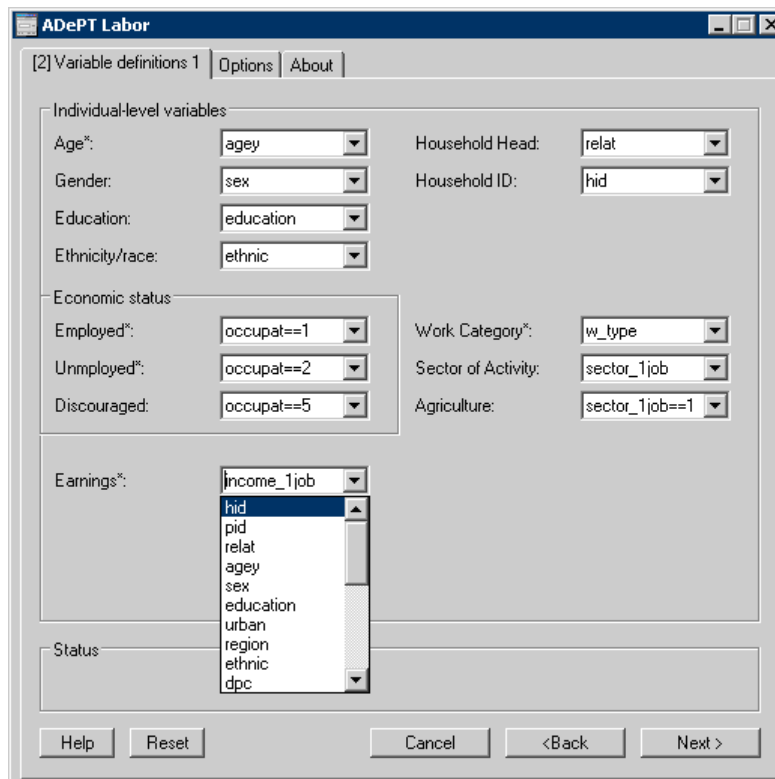
Similarly, if a file can not be created at the specified location (e.g. on a read-only media like CD or DVD disks, or in a system folder, where the system denies access) the user will get a message suggesting to provide a different location.

⁴ Provided that the output file has not been changed by any other application (even Excel)

2.3 Step 2. Variables specification

The second step is association of variables in the user's data with the meta-variables used in the program.

The required variables (marked with a star in the dialog) are necessary to build any tables at all: age, employed and unemployed dummies, work category indicator, and earnings.



The screenshot shows the 'ADePT Labor' dialog box with the following settings:

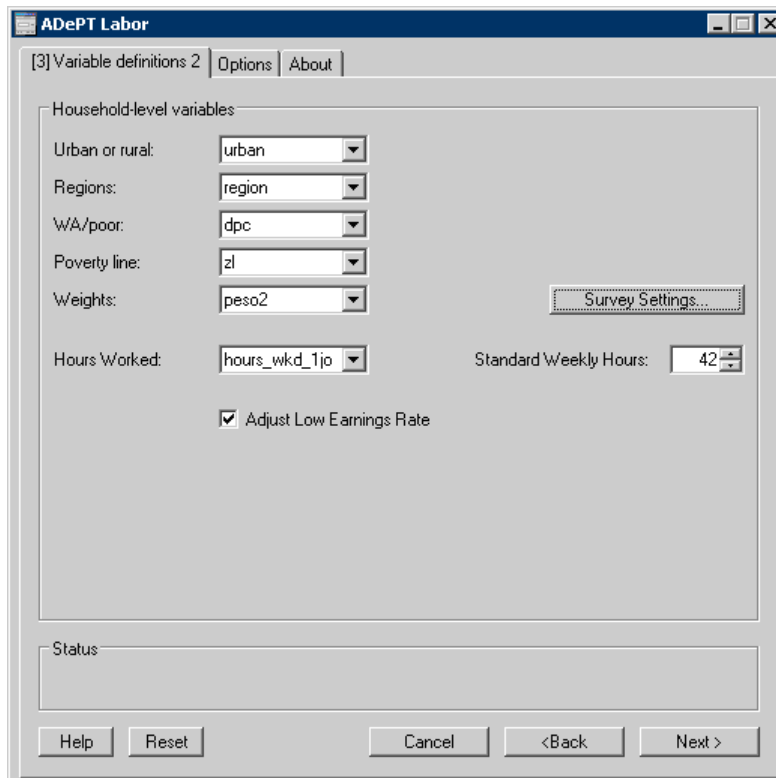
- Individual-level variables:**
 - Age*: agey
 - Gender: sex
 - Education: education
 - Ethnicity/race: ethnic
- Economic status:**
 - Employed*: occupat==1
 - Unemployed*: occupat==2
 - Discouraged: occupat==5
- Household-level variables:**
 - Household Head: relat
 - Household ID: hid
- Work and Activity:**
 - Work Category*: w_type
 - Sector of Activity: sector_1job
 - Agriculture: sector_1job==1
- Earnings*:** income_1job
- Status:** A list box containing: hid, pid, relat, agey, sex, education, urban, region, ethnic, dpc.

Buttons at the bottom: Help, Reset, Cancel, <Back, Next >

The drop-down lists contain variables from the files specified by the user at step 1. Alternatively, variable names can be entered directly into the fields of the form. The more (optional) variables are defined in the data, the more tables will ADePT be able to create.

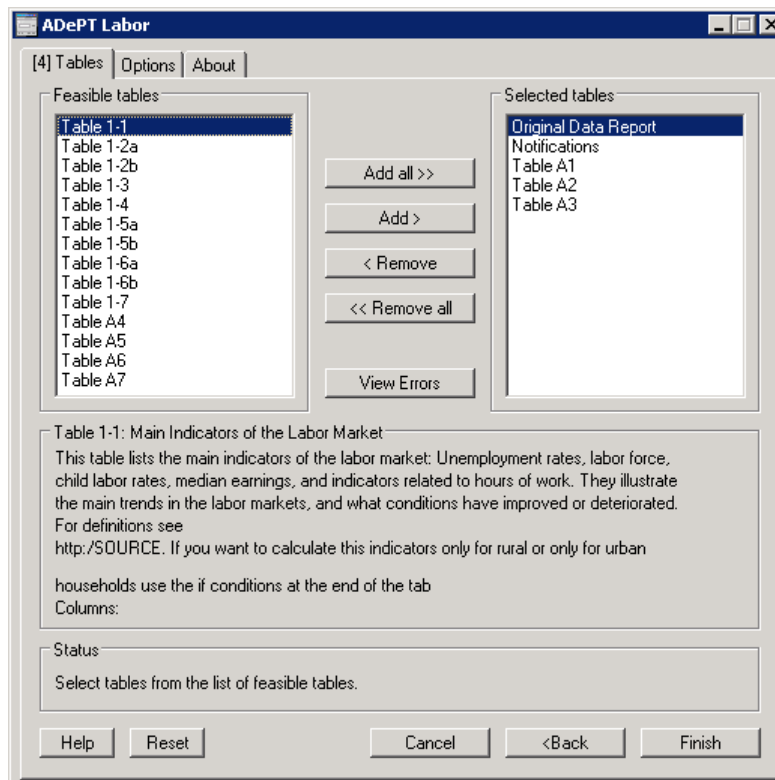
It is possible to specify expressions instead of variable names to recode the variable “on-the-fly” without modifying any data files when the coding of a variable differs from the one assumed in ADePT. In the illustration above a dummy variable for “unemployed” status will be generated internally from the categorical variable *occupat*. Only those individuals, which have *occupat* equal to 2 will have status “unemployed”. For more information on specifying expressions please see the corresponding advanced topic in this guide.

After the individual-level variables are specified, the user may provide some household-level variables (they are all optional) and the weighting variable (if present). If the survey structure is more complex, e.g. stratification must be properly accounted for, the “**Survey Settings**” button provides access to the standard Stata survey settings dialog. More information on weighting can be found in the advanced topics of this guide.



There are two additional parameters here: “*standard weekly hours*” allows to specify the standard duration of the working week (1-100 hours) and there is a possibility to adjust low earnings rate if necessary.

2.4 Step 3. Tables specification

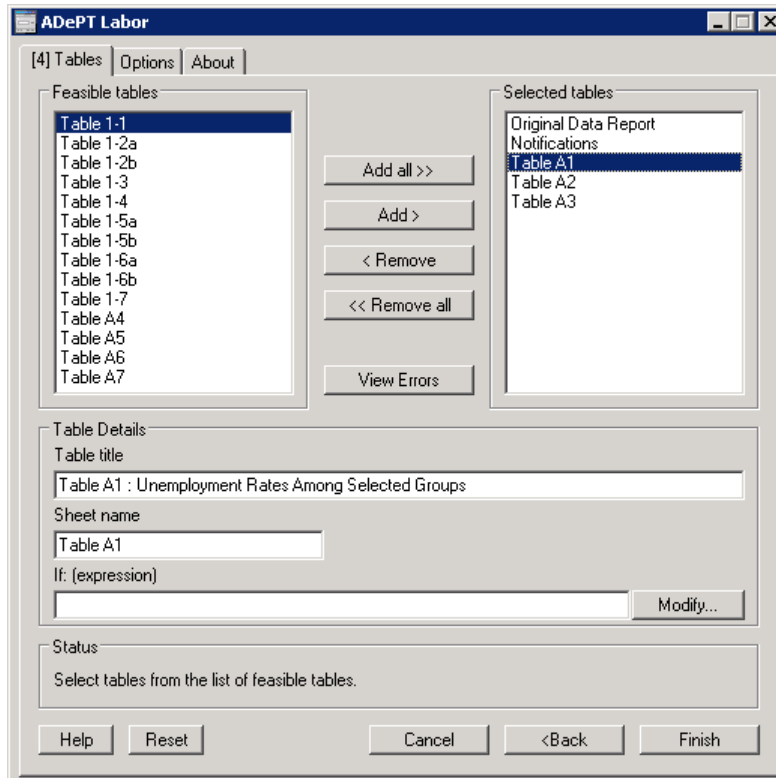


This screen informs users which tables can be built with the set of variables specified. Depending on the number of variables defined in user files, the set of feasible tables (shown in the left panel of the screen) will vary.

The four buttons located between the two panels allow moving a table between one list (feasible tables) and the other list (selected tables). Each table of the list can be moved individually ('**Add >**' and '**< Remove**') or the whole list can be moved at once ('**Add all >>**' and '**<< Remove all**').

Two tables will appear in the list of selected tables by default: "*Original Data Report*" and "*Notifications*". We recommend always having these two tables included. "*Original Data Report*" provides information on ranges, means, percentiles and numbers of unique values for each used variable from the user data. "*Notifications*" lists all the warnings and notes that ADePT program issues to the user.

A brief description for each table is provided in the lower panel of the screen. It is possible to change a title of a table by clicking on this table in the right panel. After that, the description of the table disappears and adjustment parameters for this table are shown instead, as illustrated in the following figure:



Each table will be placed into its own sheet in the spreadsheet file. The title of this sheet, as well as the title of the table can be changed in screen 4. If-conditions are discussed later in this manual in advanced topics.

2.5 Results

After '**Finish**' is clicked on screen 4, the program starts building selected tables. The title of the currently built table is displayed in the status line.

After the last table is built, ADePT attempts to open the XML output file automatically on a Windows based computer. If *Microsoft Excel* or *OpenOffice Calc* spreadsheets are not found on the user computer, ADePT will inform the user about the location of the XML file and will wait for further instructions. The generated XML file is a spreadsheet with multiple sheets, each corresponding to one table.

Summary statistics reports a standard description of used variables in all data files. The following statistics are reported: number of observations (N), average value (*mean*), minimum value (*min*), maximum value (*max*), first percentile ($p1$), 50th percentile ($p50$), 99th percentile ($p99$), and number of unique values (*unique*).

Points in the column '*unique*' indicate that there are many unique values (in other words, this variable is continuous).

Notifications sheet contains notes, warnings and error messages issued by ADePT. The color of the tab of this sheet depends on the type of the messages listed.

- If any errors are detected, the tab is colored red.
- If no errors are detected, but some warnings are present, the color of the tab is yellow.
- If neither warnings nor errors were listed, the color of the tab is default (white).

It is important to review the messages issued by ADePT in order to avoid serious mistakes in the generated tables.

If an error needs correction, close the XML file in the spreadsheet program, return to ADePT, use '**<Back**' button to return to the screen where an error was made. Change the settings as necessary and proceed with the '**Next>**' button. Sometimes errors must be corrected in the user's data files. In this case it might be useful to save the session using a corresponding button on the '*Options*' tab and resume the session later by loading the saved configuration file.

ADVANCED TOPICS

- 3.1 Expressions in variable fields
- 3.2 If-conditions for tables
- 3.3 Loading and saving program configurations
- 3.4 Automatic saving and loading of configurations
- 3.5 Specifying options to include standard errors and frequencies
- 3.6 Checking the variables
- 3.7 Survey design settings

3.1 Expressions in variable fields

Some of the variables (e.g. urban and gender) are treated in ADePT as dummy variables and thus must have only two values (if the variable is used). Internally ADePT assigns specific meaning to particular values of these variables, for example, 1 denotes urban population. However the coding of these variables might be different in the user's datasets. If this is the case:

- recode original data to match ADePT's expectations, e.g. 1=male, 1=urban, 1=household head, or
- specify an expression instead of a variable name, which ADePT will evaluate and take the result as a corresponding indicator.

Example: in the original dataset a variable *URBANPOP* takes values 1 for rural and 2 for urban population. Because there are several data files (each corresponding to a different year) it may be cumbersome to recode this variable into the format that ADePT expects. It might be easier to specify an expression of the following type:

```
URBANPOP==2
```

ADePT generates a new variable in each of the datasets specified. The variable will take the value one whenever *URBANPOP* was equal to two. For all other values (in this case ones) this new variable will take a value of zero. Missing values of *URBANPOP* will remain missing.

In general, ADePT accepts the following expressions:

```
varname  
varname==const  
varname!=const  
varname> const  
varname>=const  
varname< const  
varname<=const
```

Note that the name of a variable must be the first word of the specified expression. Whenever an expression evaluates to logical 'true' a value one (1) is taken, while zero (0) is taken for logical 'false'.



Note: ADePT never modifies any user data files. However during its work it creates temporary copies of the user data and any modifications are performed on these copies.

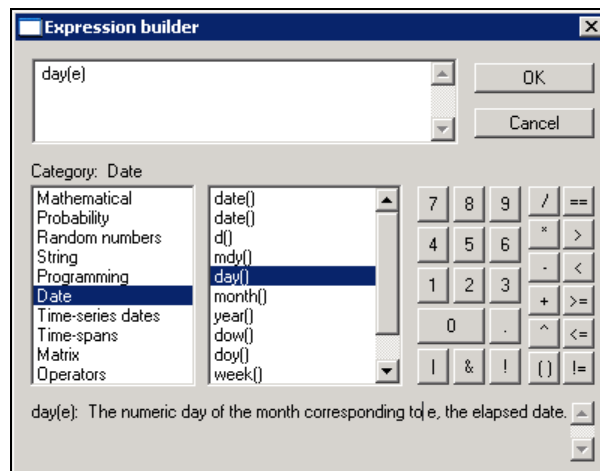
3.2 If-conditions for tables

If-conditions allow focusing on a particularly interesting group of individuals for a detailed analysis. ADePT will pass these conditions directly to Stata as so-called '*if-qualifiers*' before generating corresponding tables⁵. A standard Stata syntax may be used and it is possible to use other variables (besides the variables used in the screens 2 and 3) in the if-conditions.

Stata logically evaluates the if-condition for each observation to either '*true*' or '*false*'. Only those observations, for which the result of this evaluation is '*true*' will be used for generation of the table.

Example: suppose a user wants to implement analysis separately by indigenous status. There is a variable *INDIGENOUS* in the dataset, but it is not asked by ADePT and no special table for analysis of indigenous/non-indigenous population listed. In this case if-conditions *INDIGENOUS==1* (*INDIGENOUS==0*) may be specified for one of the standard tables to restrict the tabulation to indigenous (respectively non-indigenous) population.

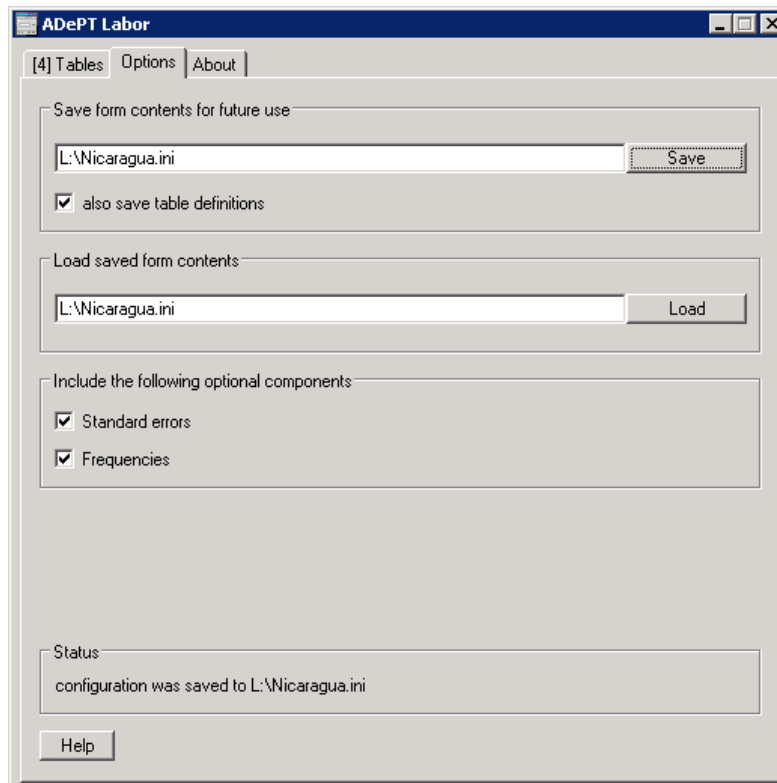
For more complex expressions, **Stata** expression builder dialog might become useful. It appears on the screen after a click on the '**Modify...**' button and allows the user to click on functions and operators with a mouse, instead of typing them. A brief reminder on the purpose and parameters of each function also appears in the dialog:



⁵ Although "Summary statistics" and "Notifications" appear as standard tables in the list, if-conditions are not applicable to them.

3.3 Loading and saving program configurations

It is possible to save ADePT's configuration for future use. To do this, switch to the '**Options**' tab by clicking on it. The '**Options**' tab will appear as shown below:



To specify a filename to save the configuration to, type its name in the top-most edit box and then click '**Save**'.

If a checkbox '*table definitions*' located under the filename is checked, the table definitions (table title, sheet title and table-specific if conditions) will be saved along with other parameters into the configuration file.



Note: The users should not modify configuration files manually.

To load a previously saved configuration, type in a filename into the second edit box on the '**Options**' tab and click '**Load**'.

After a configuration file is loaded in ADePT the name of the configuration file will be displayed in the status line and switching away from the '**Options**' tab will bring the user to the first screen (disregarding the step at which the configuration file was saved or loaded).

The complete filenames of the user's source datasets are written into the configuration file. It is important that the datasets are located in exactly the same folder(s) if this configuration file is opened

in ADePT on another computer. If the data files are missing, any configuration referring to such files may cause serious errors.

ADePT automatically determines if table definitions are present in the configuration file. If no tables are found in the configuration file – a standard set of tables will be used (a standard set of tables is also set after clicking on '**Reset**').

3.4 Automatic saving and loading of configurations

ADePT attempts to load the last successful configuration during its start-up. If this configuration is found, it will be loaded automatically and the user is notified about it in the status line. This configuration is saved to an ADePT system file after the tables are produced. The user can:

- continue using the previous configuration, or
- go to '**Options**' tab and load another configuration, or
- click '**Reset**' to empty all the fields and start from the beginning.

If no previously saved successful configuration is found, ADePT will start with empty fields.



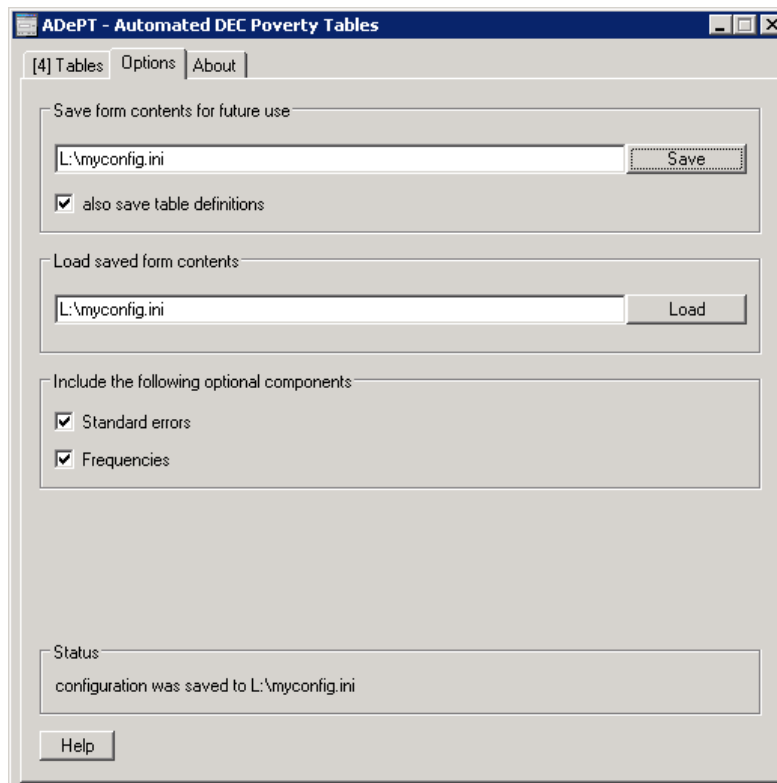
Note: changes in table titles and if-conditions are not saved automatically. Using standard Save and Load dialogs allows saving table definitions as well.

3.5 Specifying options to include standard errors and frequencies

Frequencies and standard errors can be generated on separate sheets in the output file. Large standard errors indicate imprecise estimation of the figures reported in the selected table. Low frequencies indicate that the results may not be reliable. The tables of errors and frequencies are similar to the main tables in their layout.

Both options are switched off by default.

To switch on or off generation of the tables of standard errors and frequencies, click '**Options**' tab. The two checkboxes '**Standard errors**' and '**Frequencies**' control generation of these additional output components:



3.6 Checking the variables

ADePT performs a number of checks on the user input and signals with an error message or a warning whenever it detects any problems with the data. The following table summarizes the checks conducted by ADePT:

What is checked?	What is the rule?	What happens if the check fails?
Every variable	Variable must be present in all files and have the same name in all files.	*
Every variable	Variable must be numeric	*
Categorical variables	Number of distinct values must be within an expected range	Warning and continue
Categorical variables	Codes of categories must be integer numbers	*
Categorical variables	Coding of variables must be consistent across all the datasets	Warning and continue
Categorical variables	Coding of variables must have no “suspicious” codes	Warning and continue
Dummy variables	Number of distinct values must be: 2	*
Dummy variables	Must have 1 among their values	*
Every expression	Expressions must have no syntax errors	*
Age	Number of unique values must be more than 20 Values must be in range [0;120]	*
Hours	Must be within the range [0;100]	Warning and continue
Weights	Must be non-negative	*
HouseholdID	If specified, number of unique values may not be more than 80% of the number of observations, otherwise it must be a household-level file	Warning
Household head	Number of unique values must be 2	Warning
Household ID and Household head	There should be exactly one head in each household	Warning

For the checks marked with a star (*) if the check fails:

- If the variable is required, ADePT will stop with an error message.
- If the variable is optional, ADePT will continue as if this variable was not specified.

See section 1.4 for variables classification into continuous, categorical, and dummy variables.

Some of these problems have an immediate impact on the whole process, while others cause warnings and notifications to be written in a special file, which is presented to the user together with the tables. If any error is detected in the user file, a special button **‘View errors’** will appear in the tables specification screen (step 4). A click on this button opens a file containing error messages, warnings and notifications. If no errors occurred, the button will not be present on this screen, but notes and warnings will be nevertheless written to a special sheet *‘Notifications’* in the output

spreadsheet. If the list of feasible tables does not show a table the user expects to see there, reviewing the error messages file will help finding out the reason why the table is not feasible.

The consistency check procedure relies on an assumption that all possible values of a categorical variable must be observed in each of the datasets. Whenever a variable appears to have inconsistent coding across the different years ADePT issues a warning. The warnings contain information on the name of the inconsistent variable, values which appear to be inconsistent and the files in which they occur while they are not expected to occur (or conversely the files in which they are missing when they are expected to appear). If a variable is reported being inconsistent, one of the following may be done:

- ignore the warning and do nothing, in the case when coding is as it is supposed to be, or
- if a category which makes coding of the variable inconsistent actually denotes missing (e.g. codes 999 often denote missing values in the user's datasets but are not recognized as such by Stata automatically) this category must be recoded to missing prior to running ADePT, or
- if coding of a variable has changed between the different datasets it might be possible to harmonize it by merging some categories together.

Inconsistency in coding of categorical variables may lead to the empty spaces in some of the tables, which use these variables.

AdePT will try to detect 'abnormal' values in categorical variables, e.g. if education is coded 1, 2, 3, 4, 5, 999, AdePT will issue a warning stating that 999 is suspected to be an outlier. Here is how it looks like among the other messages:

	A	B
1	[FILE]	Running variable checks in L:\Ukraine\2003_ind.dta
2	[NOTE]	region does not have any value labels defined
3	[WARNING]	Suspected outliers with code(s): 999 - in variable edlev
4	[NOTE]	ecstatus does not have any value labels defined
5	[WARNING]	Suspected outliers with code(s): 999 - in variable ecstatus

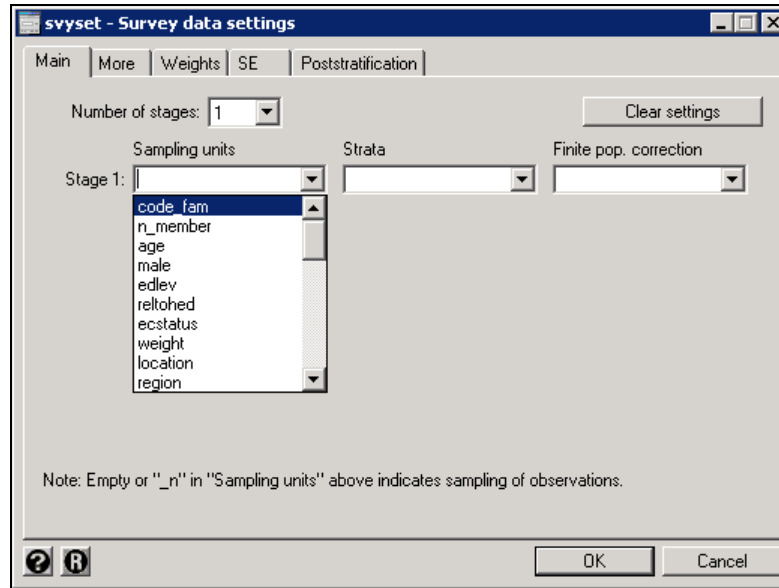
This check is implemented without binding to particular values (e.g. 9, 99, 999, etc.). It is the set of possible values that matters, e.g. 135 will also be considered an outlier if the only other values are 1, 2, 3, and 4. The algorithm (formally known as Grubbs' test for outliers) is sensitive to the gaps in coding and is calibrated in such a way that it produces warning messages if there is a sufficiently large gap between the categories (the sensitivity parameter can not be modified by the user). Thus it is a heuristic algorithm, which may only suspect, that something might be wrong with the variable coding.

ADePT checks if a particular value is among the values of the variables. E.g. ADePT verifies that the dummy variables contain 1 (one) among their values. All other non-missing values in dummy variables are automatically recoded to 0 (zero). This process can be controlled by specifying proper expressions if the value indicating the group of interest is different from 1 (see "expressions in variable fields").

ADePT checks if there is exactly one household head in every household (provided that a variable, which determines relationships in the household, is specified). A large number of errors of this type may signify a coding error of the relationship variable. These errors, however, are not critical for ADePT, they are errors in the data.

3.7 Survey design settings

ADePT exploits Stata standard facilities for handling complex survey data including multiple stage sampling and poststratification. Clicking on '**Survey Settings...**' button in the screen 3 brings the following dialog:



More details can be found in the Stata manual on survey commands or online (type `help svy` in the `Stata` command prompt).



Note: some Stata datasets come with the survey design options saved together with the data in the `.dta` files. By default ADePT ignores this information. To load the survey design settings from the data file (if contained there) click '**Survey Settings...**' button (if survey design options were saved with the data they will appear in the fields of the dialog). Click '**OK**' to accept these settings. The survey settings contained in the first file will be used for all files.